DOCUMENT RESUME

ED 254 531 TM 840 795

AUTHOR Farland, Dale S.; Gullickson, Arlen R.
TITLE Cooperative Test-taking.
PUB DATE 6 Oct 84
NOTE 17p.; Paper presented at the Annual Meeting of the Northern Rocky Mountain Educational Research Association (Jackson, WY, October 6, 1984).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; College Seniors; *Cooperation; *Grouping (Instructional Purposes); Higher Education; Learning Strategies; Scores; *Student Attitudes; Student Evaluation; Test Format; *Testing; Test Wiseness
IDENTIFIERS *Cooperative Test Taking

ABSTRACT

The effect of cooperative test-taking (CT) on the achievement and attitudes of college seniors was investigated. The students (n=46) were randomly assigned to one of two sections of a course in measurement. They worked on projects in randomly assigned groups of four or five students. One section took frequent cooperative quizzes, while the other took quizzes individually. Results indicated quiz scores were more homogeneous within the cooperative groups, and generally higher for the CT section. There was no significant difference between sections on the two regular examinations. Attitudes toward CT and beliefs about its fairness were more positive for students from the CT section. Students in this section were more likely to believe that their study groups influenced how much they learned and how well they did on quizzes. Using CT as one part of a student's evaluation may have positive benefits while yielding equal levels of achievement. Cooperative test-taking was interpreted by students as a learning opportunity, as well as an evaluation tool. Students in the CT section were tested individually on examinations as part of the total evaluation. (Author/DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED254531

Cooperative Test-taking

Dale S. Farland and Arlen R. Gullickson

School of Education

University of South Dakota

Paper presented at the Annual Meeting

of the Northern Rocky Mountain Educational Research Association

Jackson, Wyoming

October 6, 1984

TM 840 795

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D.S. Farland

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running head: COOPERATIVE TEST-TAKING

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Abstract

The study investigated the effect of cooperative test-taking (CT) on the achievement and attitudes of 46 college seniors, randomly assigned to one of two sections of a measurement course. As part of the course, the students worked on projects in randomly assigned groups of four or five. One section took frequent quizzes cooperatively; the other section took the quizzes individually. Quiz scores were more homogeneous within the cooperative groups, and quiz scores were generally higher for the CT section. There was no significant difference between sections on the two exams. Attitudes toward CT and beliefs about its fairness were more positive for students from the CT section. Students in the CT section were more likely to believe that their study groups influenced how much they learned and how well they did on quizzes. Using CT as one part of student evaluation may have positive benefits while yielding equal levels of achievement.

Cooperative Test-taking

The practices of cooperation and test-taking have not been viewed as compatible in American education. The taking of an individual test has generally been considered the standard measure of satisfactory academic performance. As such, cooperation in the completion of a test is antagonistic to the purpose of testing. Three recent, comprehensive, widely used measurement texts (Gronlund, 1981; Hopkins & Stanley, 1981; Mehrens & Lehmann, 1984) do not mention an option of students working collectively on tests, other than as inferred in discussions of cheating.

However, in recent years there has been much interest in the use of cooperative learning methods in other aspects of education. A number of authors have spoken at length about the advantages of cooperative learning (Johnson & Johnson, 1975; Johnson, Johnson, Holubec, & Roy, 1984; Johnson, Maruyama, Johnson, Nelson, & Skon, 1979; Sharan & Sharan, 1976; Slavin, 1983; Webb, 1982). The advantages of cooperative learning are outlined in detail in the references cited above, but in essence there are two major advantages. The research generally shows a positive effect on student achievement when cooperative learning methods are used, and there is also a positive effect on non-cognitive outcomes such as self-esteem and attitudes toward other students and school (Slavin, 1983).

This study addresses the assumption of incompatibility of testing and cooperation and it attempts to reconcile the two practices. By applying methods of cooperative learning to testing practices, perhaps some of the advantages of cooperation can be obtained, and some potential problems associated with testing can be avoided.

Beliefs about the necessity of individual measures of academic performance are deeply held. Even when authors write persuasively about the prevalence of cooperation in society and the need for cooperative methods in schools (Johnson & Johnson, 1975), individual accountability is still relied upon for evaluation (Johnson et al., 1984, p. 32). In a review of seven methods of implementing cooperative learning, Slavin (1983, p. 24), reports that none of the seven has as a standard procedure a collective effort on quizzes or tests. However, in several of the methods, students produce a product cooperatively. Some of the methods suggest that student quizzes, taken individually, be combined for a group grade.

We propose that individual assessment of student achievement be used, but that it be combined with cooperative test-taking opportunities. What we have assessed in this study is the opportunity during a course to work cooperatively on frequent assessment activities (quizzes) that would ordinarily be done individually. In the method used in this study, students are still required to take examinations individually. In the typical research on cooperative learning, the focus is upon the learning activities that take place prior to the assessment of achievement.

Tests are often hoped to be a learning opportunity for students, even though it is unlikely that many students view them as such. An advantage that may be obtained from cooperation on tests or quizzes is enhanced learning. If students are allowed to cooperate while taking a test, it is more likely they will think of the time as a chance to learn, and if the cooperative learning research applies, perhaps they will in fact learn from the experience. It must be pointed out that the actual purposes for

measurement have more to do with decision making than providing a learning opportunity (Mehrens & Lehmann, 1984, p. 7). Again, we do not suggest that CT be used for all assessment, at least as implemented in this study.

Anxiety is a problem that has for some time been associated with tests (Paulman & Kennelly, 1984). Motivation texts report that performance declines if arousal (anxiety) becomes too high, and that the effect is more pronounced as tasks become more complex (Weiner, 1980, p. 136). Teachers are encouraged to reduce factors that may arouse anxiety in the testing situation. The opportunity to work with others on a test may reduce the anxiety associated with the test and result in better student performance.

There are trade-offs associated with allowing students to work cooperatively on tests. The validity of an individual measure of performance is certainly compromised if other students are allowed to help during the assessment. The design of the method can compensate for this by using CT for only part of the assessment. For example if regular quizzes are given, they can be administered using CT but examinations can be done under standard conditions. The fairness of the measurement may be questioned even by students (Johnson et al., 1984), but they report that when given an opportunity to participate in cooperative incentive structures, students begin to consider cooperative rewards as fair.

The diffusion of responsibility that is possible when groups work collectively on a task may allow students to be less careful in their individual preparation, hoping to profit from others' work. In cooperative learning research, this is guarded against by building in conditions that make individual effort public. Ways to do that are to have each student

responsible for completing his or her own answer sheet, to encourage the students to actively discuss answers, and to assign students to groups to minimize the chance for friends to form their own group with ulterior motives. Further, without individual incentive to learn the content and therefore to perform well on a test, one of the chief sources of academic motivation would be eliminated. This concern is avoided to some extent by the use of examinations that are done individually.

As mentioned earlier, the testing literature seems not to have addressed the possibility of CT. The cooperative learning literature discusses the use of cooperative tasks of various kinds, typically not examinations, but we found no research that specifically addresses the effect of CT on achievement or attitude.

Therefore, this study sought answers to three questions. Do student's quiz scores vary as a function of the opportunity to work cooperatively while taking them? Is there an effect on learning (individual exam performance) if students are given the opportunity to work cooperatively on quizzes during a course? What are student's perceptions of the opportunity to take quizzes cooperatively during a course?

Method

Subjects

Forty six seniors at the University of South Dakota, during their student teaching semester, were enrolled in a five week educational measurement course. The students in the course were randomly divided into two sections of 21 and 25 students. Several students dropped the course after the beginning of the semester and another reported to the wrong

section, accounting for the uneven split. The CT condition was randomly assigned to the section that ended up having 25 students.

Procedure

The two sections met during mornings Monday through Friday each week, with the individual test-taking (IT) section meeting first, followed by the CT section. The same professor taught both sections. At the beginning of the semester, each section was randomly divided into groups of four or five students. The groups within both sections participated in discussions and spent a small part of most class periods working within their groups. The sections were given quizzes every other class day (a total of ten). The CT section was allowed to work on the quizzes in their small groups. The IT section was told that they would be quizzed in a typical classroom testing fashion, with each student working individually on his or her own quiz. The sections were given the same amount of time to complete the quizzes. Students were informed of the testing differences between sections.

The CT groups were encouraged to discuss the quiz and collectively decide upon answers. Each student chose and recorded his or her own answers on individual answer sheets. Scores were assigned individually to students, based upon the answer sheet each submitted. Observations of the CT groups during the quizzing showed heated discussion over possible answers. Quizzes counted toward the course grade, which was based upon fixed percentages of possible points rather than on a curve, thereby minimizing need for competition among students. The midterm and final exams and the attitude instrument were completed individually on the last day of class.

Quizzes each consisted of five multiple choice and true-false items. The midterm and final exams each had about 80 items, parallel in content and format to those on the quizzes. The final exam included items covering the first half of the course. Both exams included about five items drawn from the quizzes; students were told that some items would be repeated.

Results

Multivariate analysis of variance (SPSS-X MANOVA) was performed on the achievement and on the attitude data. The factor of interest in the analyses was testing condition (cooperative versus individual). The first analysis used the ten quizzes as dependent variables. A second analysis included the midterm and final exams as dependent variables. Two separate analyses of achievement data were performed because of the difference in administration procedures used for the two types of exams (all students took the midterm and final individually). Responses to items on the attitude instrument were the dependent measures in the analysis of the attitude data.

Eight of the students had missing data on one or more of the quizzes, leaving 38 cases for the analysis of the quiz performance. Box's multivariate test for homogeneity of dispersion matrices was significant $F(55, 3797) = 2.21$, $p < .001$. A significant Bartlett test of sphericity (45) = 111.55, $p < .001$ as was $F(max)$ (10,36) = 3.49, $p < .05$. Because of the failure to meet assumptions for the univariate approach to the analysis, the multivariate tests are considered. For all quizzes, the CT sections had a smaller standard deviation of quiz scores than the IT section (see Table 1). The analysis of the quiz scores yielded a significant multivariate test of significance (Wilks), $F(10, 27) = 5.59$, $p < .001$. However, only the

univariate test for the second quiz was significant, $F(1,36) = 12.13$, $p < .01$. Three other quizzes showed sizeable but not significant differences between the two sections. However, the quiz means did not consistently favor either section (see Table 1). On half of the quizzes, the IT section had a higher average than the CT section.

All 46 cases had complete data for the analysis of the exam scores. Again, Box's multivariate test for homogeneity of dispersion matrices was significant $F(3, 6935351) = 4.11$, $p < .01$, as was Bartlett's test of sphericity $(1) = 15.10$, $p < .001$. The $F(max)$ $(2, 44) = 1.58$, $p > .05$ was not significant. The CT section had a larger standard deviation on the midterm exam and a smaller standard deviation on the final exam (see Table 2). The multivariate test of significance (Wilks) was not significant, $F(2, 43) = .21$, $p > .81$ (see Table 2 for the means).

Complete data on the attitude instrument was available for all 46 cases. Box's multivariate test for homogeneity of dispersion matrices was significant $F(210, 5544) = 1.23$, $p < .05$, as was Bartlett's test of sphericity $(190) = 305.88$, $p < .001$. The $F(max)$ $(20, 44) = 3.20$, $p > .05$ was not significant. The Wilks multivariate F for the testing condition main effect over the 20 attitude items was significant $F(20, 25) = 2.08$, $p < .05$. The univariate analyses indicated that the students in the two testing conditions differed in their response to four of the items. Those items and the corresponding statistics are reported in Table 3.

The IT students were more inclined than the CT students to want to take quizzes individually (item 1). The CT students felt that their quiz performance was enhanced by their group (item 2) and that it was fair to

have grades determined in part by group performance (item 6). The IT students believed that they learned the material better because of being required to take quizzes individually.

Discussion

The premise of this research was to investigate the potential of using cooperative test-taking (CT), a seemingly illogical combination. The advantages suggested in the introduction were that students would learn more, express more positive attitudes, and would experience less anxiety under conditions of CT. The potential disadvantages were that the validity of individual scores would be compromised and that students would take advantage of the opportunity to not prepare for quizzes.

The results of the study are positive or at least neutral on all counts. The examinations were traditional measures of individual performance in the course, and there was no difference on the exams between the CT and IT sections. There was therefore no evidence for superior learning on the part of the CT students, nor was there evidence for the CT students to not study the material because of diffusion of responsibility. One possibility not controlled for was the amount of "cramming" done by either section prior to the exams. If students in either section crammed, they could hide any effect on the exams that was due to the treatment.

The CT students did better on the quizzes than the IT students. The cooperative learning research (Slavin, 1983) indicates that when given an opportunity to share resources (in this case, information about the subject matter), collective performance is better than individual performance. Slavin states that "many [of the studies] simply found that two or more

students who take a test together do better than students who work alone" (1983, p.12). It is surprising that there was a significant difference on only one of the quizzes, and that was early in the course (quiz 2).

There may be some difference in the effect of cooperation when comparing a typical educational test or quiz to the performances involved in the cooperative learning literature. The type of task used in earlier research involves building block towers and other activities not a traditional part of education. Those activities usually have a specific and obvious right answer, once that answer is suggested. That was clearly not the case on the quizzes, as students were overheard to complain about letting others talk them out of an answer that would have been correct.

The CT students indicated positive attitudes about the use of CT, about the influence of their groups on their quiz scores and learning, and the fairness of basing grades in part on group efforts. There were no differences on attitude items assessing general attitudes toward the course, nor on anxiety associated with the quizzes.

Questions remain about the real effects of CT upon the study habits of students. Slavin (1983) found in his review that there was more "helping" among students under a cooperative incentive structure than under either individual or competitive incentives. The specific result of helping is not always clear. Hamblin, Hathaway, & Wodarski (1971) do show a correspondence between helping in prior studying and quiz performance (on individual quizzes), but they did not specifically manipulate helping.

Perhaps a more detailed look at what is learned during the course, in preparation for the quizzes, would indicate differences. Other

manipulations of the reward structures and the task structures are possible and should be explored in an effort to maximize the benefit of CT. For example, under the quizzing procedures used in the CT groups, it was possible for the diffusion of responsibility to be high (a student could substitute for another in providing information about each item). There may have been less pressure on students to prepare for quizzes as well as they might have. We have additional research currently underway to further explore the specific effect of CT on studying and quiz performance.

References

Gronlund, N. E. (1981). Measurement and evaluation in teaching (4th ed.). New York: Macmillan.

Hamblin, R. L., Hathaway, C., & Wodarski, J. (1971). In E. A. Ramp & B. L. Hopkins, A new direction for education: Behavior analysis 1971, volume one. Lawrence, KS: University of Kansas.

Hopkins, K. D., & Stanley, J. C. (1981). Educational and psychological measurement and evaluation (6th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Johnson, D. W., & Johnson, R. T. (1975). Learning together and alone: Cooperation, competition, and individualization. Englewood Cliffs, NJ: Prentice-Hall.

Johnson, D. W., Johnson, R. T., Holubec, E. J., & Roy, P. (1984). Circles of learning: Cooperation in the classroom. Alexandria, VA: ASCD.

Johnson, D. W., Maruyama, G., Johnson, R. T., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. Psychological Bulletin,

89, 47-62.

Mehrens, W. A., & Lehmann, I. J. (1984). Measurement and evaluation in education and psychology (3th ed.). New York: Holt Rinehart & Winston.

Paulman, R. G., & Kennelly, Kevin J. (1984). Test anxiety and ineffective test taking: Differnet names, same construct? Journal of Educational Psychology, 76, 279-288.

Sharan, S., & Sharan, Y. (1976). Small-group teaching. Englewood Cliffs, NJ: Educational Technology Publications.

Slavin, R. E. (1983). Cooperative learning. New York: Longman.

Webb, N. M. (1982). Student interaction and learning in small groups. Review of Educational Research, 52, 421-445.

Weiner, B. (1980). Human motivation. New York: Holt, Rinehart and Winston.

Table 1

Mean Quiz Scores for the Cooperative Test-taking (CT) and for the Individual Test-taking Conditions (IT)

	Testing condition			
	IT (n = 17)		CT (n = 21)	
Quiz	M	SD	M	SD
1	4.47	0.80	4.33	0.58
2	3.88	1.05	4.76	0.44
3	4.00	0.94	4.43	0.51
4	3.29	1.45	4.05	1.12
5	3.76	1.09	3.57	0.51
6	3.06	1.09	2.95	0.74
7	3.00	1.50	3.14	0.73
8	3.71	0.85	3.48	0.51
9	4.47	1.07	4.24	0.62
10	2.76	0.90	3.24	0.62

Table 2

Mean Test Scores for the Cooperative Test-taking (CT) and for the Individual Test-taking Conditions (IT)

	Testing condition			
	IT (n = 21)		CT (n = 25)	
Test	M	SD	M	SD
1	54.29	6.72	53.20	11.70
2	49.57	9.00	48.08	6.55

Table 3

Mean Attitude Item Rating and F Values for the Significant Items for the Cooperative Test-taking (CT) and for the Individual Test-taking Conditions (IT)

		Testing condition				
		IT (n = 21)		CT (n = 25)		
	Item	M	SD	M	SD	F
1	choose to take quizzes individually	4.19	1.57	2.16	1.31	23.85**
2	quizzes better because of my group	2.76	1.67	4.68	1.14	21.18**
6	fair to base grade in part on others in group	2.29	1.27	3.44	1.23	9.78*
16	learn better if required to take quiz alone	4.24	1.04	3.12	0.93	14.79**

Note. All rating scales were from strongly disagree (1) to strongly agree (6).

* $p < .01$. ** $p < .001$.